INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION ᴛᴿᴇᴬᴛᴵ (PCT)

(51) International Patent Classification 6 : G01N 27/26	A1	(11) International Publication Number: **WO 99/04251**
		(43) International Publication Date: 28 January 1999 (28.01.99)

(21) International Application Number: PCT/US98/14740

(22) International Filing Date: 21 July 1998 (21.07.98)

(30) Priority Data:
08/895,247   21 July 1997 (21.07.97)   US

(71) Applicant (for all designated States except US): BIO–RAD LABORATORIES, INC. [US/US]; 1000 Alfred Nobel Drive, Hercules, CA 94547 (US).

(72) Inventors; and
(75) Inventors/Applicants (for US only): BINDER, Steven, R. [US/US]; 2506 Hawthorne Terrace, Berkeley, CA 94708 (US). SCOLARI, Caroline [US/US]; 722 Johnson Street #3, Albany, CA 94706 (US). LIKUSKI, Robert, K. [CA/US]; 4430 School Way, Castro Valley, CA 94546 (US).

(74) Agents: SNYDER, Joseph, R. et al.; Townsend and Townsend and Crew LLP, 8th floor, Two Embarcadero Center, San Francisco, CA 94111 (US).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, HR, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: COMPUTER–DIRECTED DETECTION OF PARAPROTEINS

(57) Abstract

Identification and quantification of paraproteins in a sample is achieved by Fourier analysis of mobility–based electropherograms obtained from capillary electrophoresis. The use of a computer algorithm to analyze capillary electrophoresis data, provides the clinician with methods of detecting levels of paraproteins in serum as low as 0.05 g/dL. Additionally, an individual paraprotein can be located on an electropherogram and used to monitor its increased or decreased production in an individual.

COMPUTER-DIRECTED DETECTION OF PARAPROTEINS

5   BACKGROUND OF THE INVENTION

Serum protein analysis is useful for the diagnosis of lymphoproliferative
disease. Gel electrophoresis has been the traditional method of choice for such analysis.
Capillary electrophoresis (CE) has also been applied to serum protein analysis. Capillary
electrophoresis encompasses a large variety of separation modes, many of which provide
10   separations that equal or exceed the quality of separations that can be performed on slab
gels. One form of capillary electrophoresis that is particularly suitable and convenient for
many separations is capillary zone electrophoresis (CZE). Other forms include capillary
isoelectric focusing, capillary gel electrophoresis, capillary isotachophoresis, micellar
electrokinetic chromatography, and capillary electroosmotic chromatography. Separations
15   can be performed in relatively short periods of time by using high voltages, since the small
diameter and thin wall of a capillary provide efficient removal of the joule heat generated
by the voltage. Detection methods for CE include UV absorbance, fluorescence,
electrochemical, Raman, and mass spectrometric.

In gel electrophoresis, the stained gel generally contains a pattern consisting
20   of a series of dark bands on a light background. This gel response can be scanned to find
the absorbance at a series of x positions on the gel, creating a potentially quantifiable
response. This resulting response is called a densitometer trace.

A response with an x-axis of migration time is typically the data obtained
from the capillary electrophoresis experiment. The migration time axis can be changed to
25   normalized mobility by: 1) taking the reciprocal of migration time, 2) multiplying by an

2

appropriate constant, 3) zero correcting by subtracting the electroosmotic velocity, 4) dividing by the zero corrected mobility of a charged marker, and 5) multiplying by a constant, preferably -1. This process gives a more stable x-axis, normalized mobility, allowing far more precise identification and quantitation of the electropherogram

5   components. The appearance of this normalized data set also more closely resembles that of the analogous gel electrophoresis densitometer trace, a shape that is familiar to clinicians.

The gel densitometric or capillary electrophoretic response obtained from a normal human serum proteins sample is a series of relatively broad bands, due to the

10   heterogeneity of the proteins species. Paraproteins, which are homogeneous, will typically be located in a very narrow, dense band on the gel. This narrow band is referred to as a "zone of restricted mobility." On the densitometric scan or in a CE experiment, a paraprotein will show up as a sharp peak in the midst of a broader response, while in the stained gel, the paraprotein will show up as a narrow dark band within a broader one.

15   Visual inspection of the gel electrophoretic response pattern is a very subjective, technique-dependent, and labor-intensive method to detect paraproteins, but has typically been the method of choice for gel electrophoresis. The densitometric trace may also be examined for paraproteins, but this method has proven less sensitive than direct visual inspection of the gel.

20   Monitoring paraprotein production is necessary to detect and treat diseases, such as multiple myeloma. An individual suffering from multiple myeloma will produce one or more abnormal immunoglobulins in large amounts. The abnormal protein is a paraprotein which, if detected at an early stage, allows an aggressive treatment plan to be employed. Left undetected, a more extreme therapy may be required. Thus, it is

25   important to properly detect paraproteins at as low a level as possible. Large levels are easily detected with the naked eye on either a slab gel or a capillary electrophoresis electropherogram, but low level production of paraproteins are not as easily detected. An automated analysis for paraproteins in capillary electrophoresis has increased sensitivity over the densitometric response, while requiring less time and expertise than direct

30   inspection of gels.

Prior to the discoveries underlying the present invention, methods of paraprotein analysis have simply indicated the presence of paraproteins as a plus/minus. A

method that also reports suspected position(s) of paraproteins will allow easier detection, identification, and monitoring of the levels of paraprotein production in an individual.

SUMMARY OF THE INVENTION

5   It has now been discovered that paraproteins can be detected (identified and quantified) in a serum sample using a Fourier analysis of a capillary electrophoresis data set. The methods of the present invention utilize a software algorithm employing Fourier analysis to detect the presence of a zone of restricted mobility and return an estimate of the position of suspect features. The methods described herein are capable of detecting paraproteins at concentrations as low as 0.05 to 0.1 g/dL. This concentration is

10   considered clinically significant, but can be easily missed by visual inspection of an electropherogram which has not been processed by the methods herein.

In one aspect, the present invention provides methods of detecting the presence or absence of paraproteins in a serum protein sample, by;

(a)   separating components in a serum protein sample using capillary

15   electrophoresis to generate a first data set;

(b)   subjecting at least a portion of the first data set to Fourier Analysis to generate a forward-transformed data set;

(c)   selecting any forward-transformed data sets having high frequency components above a first preselected threshold;

20   (d)   filtering and back-transforming data sets selected in step (c) to provide filtered, back-transformed data sets;

(e)   identifying the magnitude and location of residual maxima in the filtered, back-transformed data sets; and

(f)   comparing the location of any residual maxima having a magnitude above a

25   second preselected threshold to a corresponding location in the first data set to detect the presence of paraproteins in the protein serum sample.

Preferably, the first data set is normalized mobility data. Employing the normalized mobility data and Fourier analysis, the location and a threshold concentration of paraproteins in a serum sample can now be consistently determined by computer

30   algorithm. Thus, the method of this invention provides more information than was

4

previously available and by more consistent methods than previously used.

These and other features, benefits and advantages of the invention are explained in further detail below.

BRIEF DESCRIPTION OF THE DRAWINGS

5   Fig. 1 is a set of four graphs which provide an illustration of serum protein analysis and the difficulty of detecting the presence of paraproteins. The paraproteins of interest in this invention are typically in the gamma region, but can occur in the beta region also. The gamma region is that region from approximately +0.05 to -0.25 on the x-axis, which represents normalized mobility. The beta region is that region of the x-axis

10   extending from approximately -0.25 to -0.4. **Fig. 1A** is a graph of a known sample containing 2.2 g/dL of IgG. **Fig. 1B** shows the same paraprotein at a concentration of 0.1 g/dL. **Fig. 1C** illustrates a sample containing 7.1 g/dL of a different IgG paraprotein. **Fig. 1D** shows a lower concentration (0.1 g/dL) of the same IgG paraprotein.

       FIG. 2 is a set of four graphs which provide an illustration of the data sets

15   produced and used in analyzing a normal serum protein sample by the method of the present invention. In **Fig. 2A, 2C,** and **2D** the y-axis is normalized absorbance and the x-axis is normalized mobility. In **Fig 2B**, the y-axis is area proportion and the x-axis is normalized frequency units. **FIG. 2A** represents the final result showing no paraprotein regions. **FIG. 2B** represents a forward transformed data set. **FIG. 2C** represents the

20   back-transform of the data in **Fig. 2B** after it has been filtered to exclude the low frequency part. **FIG. 2D** is a comparison between the original data and a back-transform of the low frequency part that was excluded in **Fig. 2C**. Although not apparent for the y-axis scale used in **Fig.2D**, the difference between the two plots in **Fig. 2D** will yield the plot in **Fig. 2C**.

25   FIG. 3 is a set of four graphs similar to those in **FIG. 2** which have been derived from data sets for an abnormal serum protein sample containing paraproteins. In **Fig. 3A, 3C,** and **3D** the y-axis is normalized absorbance and the x-axis is normalized mobility. In **Fig 3B**, the y-axis is area proportion and the x-axis is normalized frequency units. **FIG. 3A** represents the final result showing asterisks where paraprotein regions are

5

located. **FIG. 3B** represents a forward transformed data set which has been plotted as a power spectrum. **FIG. 3C** represents the back-transform of the data in **Fig. 3B** after it has been filtered to exclude the low frequency part. **FIG. 3D** is a comparison between the original data and a back-transform of the low frequency part that was excluded in **Fig. 3C**.

5   The difference between the two plots in **Fig. 3D** will yield the plot in **Fig. 3C**.

DETAILED DESCRIPTION OF THE INVENTION
AND PREFERRED EMBODIMENTS

General

The detection and quantitation of paraproteins is used for the detection of

10   multiple myeloma, and in evaluating the course of the disease and response to treatment. The visual detection of low levels of paraproteins provides difficulties for the person unskilled in the examination of electropherograms. This is best illustrated with reference to **Fig. 1**.

Fig. 1 is a set of four graphs which provide an illustration of serum protein

15   analysis and the difficulty of detecting the presence of paraproteins. The paraproteins of interest in this invention are typically in the gamma region, but can occur in the beta region also. In the examples shown in **Fig. 1** , the gamma region extends from +0.05 to -0.25 normalized mobility units. The beta region extends from -0.25 to -0.4 normalized mobility units. **Fig. 1A** illustrates a graph of a known sample with 2.2 g/dL of IgG. The

20   large paraprotein peak in the gamma region of the spectrum can be easily detected by visual examination. **Fig. 1B** shows the same paraprotein diluted to a concentration of 0.1 g/dL. The gamma region peak at this low concentration is not easily distinguished visually. Similarly, **Fig. 1C** illustrates a high concentration sample of a different IgG paraprotein. The paraprotein peak at about -0.15 on the x-axis can be easily detected by

25   visual inspection. **Fig. 1D** shows a lower concentration (0.1 g/dL) of the same IgG paraprotein which is again difficult to detect visually. As the Figures indicate, the consistent visual detection of low levels of paraproteins can be difficult, at best.

The automated detection of paraproteins is facilitated by the fact that paraprotein responses result from the manufacture of a larger than normal amount of a

6

single species in response to a tumor, leading to a sharp response. The normal serum proteins response is the result of a heterogeneous mix of proteins, which leads to a smooth response. Thus, the homogeneous component produces a locally sharper response in the normally smooth serum proteins response.

5   Thus, differences in the frequency characteristics between the two types of response may be used for purposes of detection. Sharper peaks are expected to contain more high frequency components than smooth peaks. The frequency characteristics of a selected serum proteins region of interest (usually the gamma region), is examined using Fourier analysis. The signal of interest is interpolated to provide 2^N equally spaced data

10   points, a fast Fourier transform of the data is taken, and the power spectrum is constructed by multiplying the forward transform of the signal by its complex conjugate. The proportion of signal in the power spectrum occurring in a defined high frequency range is calculated. This calculation can involve different weightings of values over different frequency regions. If the proportion of high frequency signal exceeds threshold, the

15   possibility exists that a high-frequency component (paraprotein) is present.

          If an amount of high frequency signal above threshold is found, then an additional step is carried out to ascertain the location of the high frequency region(s) in the scan. The forward transformed data is separated into two parts: a high frequency part and a low frequency part. Back-transforming the low frequency part gives a smoothed data set

20   which can be subtracted from the original (first) data set to provide a residual data set. Back-transforming the high frequency part provides the residual data set directly. Residual segments are defined and examined, and the maximum height of each residual segment is found. If the maximum positive deviation of the residual segment exceeds threshold, this location of maximum deviation is stored as a possible site of paraprotein. This step has

25   two purposes. First, it weeds out some false positives found upon examining the power spectrum alone. Second, this step gives an estimate of the location of possible high frequency (paraprotein) features.

          However, some false positives can make it through both of these steps. For example, a sudden change in shape at the end of a delimited region, due to an improperly

30   placed delimiter, can contribute high frequency components to the power spectrum, and

7

produce fairly large residuals at the ends of the delimited region under examination. To prevent these end segments from triggering false positives, a verification of the residual results is performed.

The verification of results found by Fourier analysis is accomplished using a
5   feature pick algorithm used in CDM 2.0 software (available from Bio-Rad Laboratories). A paraprotein response is expected to appear as either a crest or a shoulder. Thus, shoulders and crests found by the feature pick algorithm in the time domain are valid features of interest. This time domain information is also available in normalized mobility units. The location of the valid feature(s) found in the time domain is checked versus the
10 location of the valid residual deviation(s) found through Fourier analysis. If a feature found by this independent check matches the location of a found residual segment maximum within a specified threshold (0.05 normalized mobility units, for example), a paraprotein is considered detected, and the x-location of the paraprotein is taken to be the location of the feature(s) found by the peak-pick algorithm.

15   Once the location of paraprotein features are found, the response may be quantified, using either manual delimiting of the area under the response, or by more automated means.

The present invention provides a method of detecting paraproteins using a computer algorithm. The algorithm removes the need to visually inspect difficult spectra
20 such as those in Fig. 1B and 1D for paraprotein peaks.

Embodiments of the Invention

In one aspect, the present invention provides methods of detecting the presence or absence of paraproteins in a serum protein sample, by;

(a)   separating components in a serum protein sample using capillary
25 electrophoresis to generate a first data set;

(b)   subjecting at least a portion of the first data set to Fourier Analysis to generate a forward-transformed data set;

(c)   selecting any forward-transformed data sets having high frequency components above a first preselected threshold;

SUBSTITUTE SHEET (RULE 26)

8

(d)   filtering and back-transforming data sets selected in step (c) to provide filtered, back-transformed data sets;

(e)   identifying the magnitude and location of residual maxima in the filtered, back-transformed data sets; and

5   (f)   comparing the location of any residual maxima having a magnitude above a second preselected threshold to a corresponding location in the first data set to detect the presence of paraproteins in the protein serum sample.

The components in a serum protein sample are readily separated by capillary electrophoresis using methods well known to those skilled in the art. Capillary

10   electrophoresis facilitates the analysis of small samples using high voltages and relatively short separation times. A preferred form of capillary electrophoresis is "open capillary zone electrophoresis," in which the separation medium is a buffer solution.

Protein separations are readily performed using capillary electrophoresis with conventional capillary electrophoresis units and materials which are commercially

15   available from suppliers such as Bio-Rad Laboratories (Hercules, California, USA). Operating conditions and procedures used for the separations are similarly conventional and can be selected and employed using methods known to those of skill in the art. A particularly preferred system for capillary electrophoresis is the Biofocus 2000 with CDM 2.0 software, available from Bio-Rad Laboratories.

20   The capillaries used for serum protein separation will typically be capillaries of silica containing material, preferably fused silica whose internal surface has not been coated. Other useful capillaries are glass or quartz. The internal diameters of the capillaries will typically be from about 20 μm to about 75 μm. Preferably, the capillaries used are those having internal diameters of about 25μm to about 35 μm, more preferably

25   about 25 μm. Capillaries of the type noted and preferred for serum protein electrophoresis may be obtained commercially from Bio-Rad. In other embodiments, the present invention will use electrophoretic separations performed in slab-shaped cells and other non-capillary systems.

As noted above, separations of serum components will typically use

30   conditions which are readily determined by those of skill in the art. Preferred conditions are described in co-pending application USSN 08/610,105, filed February 29, 1996 (Attorney Docket No. 2558B-563), incorporated herein by reference. The run buffer will

9

typically be an aqueous solution of glycine with added acid or base. In one group of
embodiments, a preferred run buffer is Bio-Rad part #194-5055. Typically, the serum
samples are diluted into an aqueous buffer prior to injection into the capillary. The diluent
may be the run buffer, or it may be a lower conductivity solution to provide a higher
5   resolution through a process known as stacking. The diluent may also contain internal
standard(s) for calibrating the y-axis or markers for calibrating the x-axis. Hippuric acid
or xanthine may be used as an internal standard, as a marker, or both. The preferred
sample diluent is Bio-Rad part #194-5056.

After introduction of the diluted sample into the capillary, a voltage is
10   applied and the sample is separated into its various components. For capillary systems,
separations will be carried out using voltages of about 1 kV to about 30 kV, preferably
about 5 kV to about 15 kV. The components are resolved into bands, which migrate along
the capillary and past the detector.

Detection of the bands of proteins can be achieved by any method that is
15   known to be applicable to capillary electrophoresis. One type of detection is ultraviolet
absorbance detection. Direct UV-absorbance detection can be achieved by passing a UV
beam through the capillary, transverse to the capillary axis, and continuously monitoring
the intensity of the beam emerging after having been interrupted by solute zones migrating
across its path.

20   Detection of the component species (e.g., proteins in a serum sample)
provides a first data set. In some embodiments, the data can be used to plot an
electropherogram. Alternatively, the first data set can be subjected to further analysis to
generate an electropherogram capable of computer manipulation for area and peak height
determination, normalization and zero correction.

25   After detection, the data is generally transferred to a digital processor (often
a personal computer) as a series of digital amplitudes. The first amplitude is generally
given an index number of zero or one. The index is generally incremented by one for
each subsequent amplitude. For further analysis and/or presentation, the indices are
converted to a more appropriate quantity with corresponding units, e.g., migration time in
30   minutes.

A first data set obtained by the methods and of the type described above is
preferably mobility zero corrected and normalized according to the methods in copending

10

application USSN _____, filed May 30, 1997 (Attorney Docket Number 2558B-590), incorporated herein by reference.

The first data set, which is preferably zero corrected and converted to normalized mobility, is subjected to a Fourier analysis (including Fourier transformation of the data) or another mathematically equivalent method to provide a forward-transformed data set. Fourier transformation is a well-known mathematical process for the conversion of time or position data into frequency data. All of the first data set can be transformed at this point or just that portion which represents a paraprotein region. In a delimited data set, the limits are typically set at 0 to -0.4 and more preferably at 0 to -0.3. The delimited portion will typically correspond to the gamma region, but could also correspond to the beta region. The forward-transformed data set thus generated can be used to construct a power spectrum for visual examination.

To determine whether the distribution of the power spectrum response over frequency indicates the presence of a paraprotein, a high frequency region and a low frequency region are defined. The boundaries between the low frequency region and the high frequency region can be defined as an abrupt transition or a gradual transition. It has been found that a linear transition is both convenient and suitable. Because of noise, detection is improved by limiting the extent of the high frequency region examined. This upper boundary may also be gradual or abrupt. If the high frequencies have already been excluded by prior filtering, the results are not sensitive to the upper boundary. By altering the transition points, all high frequencies can be examined, or a smaller subset. Typically, a lower transition from 0.002 to 0.004 normalized frequency units, and an upper transition from 0.014 to 0.016 normalized frequency units are used. Those components above the preselected threshold are selected and labeled as possible abnormal regions. The preselected area threshold will typically be set at about 1×10^{-4} to about 1×10^{-2}, and more preferably at about 1×10^{-3}. The thresholds are typically set to levels corresponding to concentrations of about 0.05 to about 0.1 g/dL.

The forward transform data set is next filtered and back-transformed. Back-transformation is the reverse operation of Fourier transform. For example, if Fourier transforms are used to convert data from the time domain to the frequency domain, then a back-transform will convert the data from the frequency domain back into time. If the data has not been filtered, the back-transform will restore the original data set.

SUBSTITUTE SHEET (RULE 26)

11

Filtering emphasizes those frequency components of interest. The forward transformed data is multiplied by a function designed to keep those frequencies of interest, and de-emphasize those frequencies not of interest. Commonly, filtering is used to separate the data into high frequency parts and low frequency parts. The transition

5   between a low frequency region and high frequency region may be gradual or abrupt. Filters using a linear ramp in the transition region are convenient and suitable for this purpose. Filters typically used include low pass filters, high pass filters, bandpass filters, notch filters, or combinations thereof. Preferably, the filters used are high and low pass filters. In some embodiments, the filter used is a ramp smoothing filter. In other

10 embodiments, a square smoothing filter is used.

If a filter is applied and the data set is then back-transformed, the back-transformed data will no longer match exactly the original data set. If a low pass filter has been applied, the back-transformed data set will be smoothed. If a high pass filter has been applied, the back-transformed data set will be a residuals data set. The same

15 residuals data set may also be constructed by subtraction of the smoothed data set from the first (original) data set.

The Fourier transforms and filtering functions can be done using the appropriate software. One example is the MatLab routine named FFT from the MatLab programming environment (Math Works, Inc., Natick Massachusetts, USA).

20   The filtered, back-transformed data set can then be examined for the presence (magnitude and location) of residual maxima. These residual maxima correspond to potential paraprotein sites. Any residual maxima having a magnitude above a pre-selected threshold is considered to be the potential site of a paraprotein. The amplitude threshold is typically set at 1×10^{-5} to 1×10^{-2} normalized AU, and more preferably at

25 about 1×10^{-4} normalized AU. The amplitude threshold is typically set to levels corresponding to concentrations of about 0.05 to about 0.1 g/dL.

The data can be analyzed under various filter conditions, and if paraproteins are found using specified filtering conditions, the area can be identified as a paraprotein region. To verify the assignment, the results of an independent feature-pick routine are

30 used to confirm that a feature exists in the predicted location. An example of this independent feature pick routine is the peak detection algorithm in CDM 2.0 software, available from Bio-Rad Laboratories which selects valid shoulders and crests in the time

12

domain. This time domain information is also available in normalized mobility units. The location of these features found in the time domain is checked versus the location of the valid residual deviation(s) found through Fourier analysis. If a feature found by this independent check matches the location of a found residual segment maximum within a

5   specified threshold (0.05 normalized mobility units, for example), a paraprotein is considered detected, and the x-location of the paraprotein is taken to be the location of the feature(s) found by the peak-pick algorithm.

In this manner, through a combination of power spectrum high frequency region examination, construction and examination of residual maxima, and verification

10   using an independent feature-pick routine, the detection of paraproteins above a threshold concentration may be accomplished and the location determined. These paraproteins can then be quantified manually or automatically and monitored for change in an individual.

A pictorial representation of the present method is provided in **Figs. 2** and **3** for a normal and an abnormal serum sample. Each of the figures is a compilation of four

15   graphs which represent delimited data sets, power spectra, residual spectra and a comparison of a filtered, back-transformed data set to the delimited data set. **Fig. 2** provides the graphs for data from a normal serum protein sample and **Fig. 3** provides the graphs for data from a serum protein sample containing paraproteins. **Figs. 2B** and **3B** depict power spectra constructed from Fourier transformation of the first data sets

20   generated by the capillary electrophoresis serum protein separation. The data is plotted as normalized response versus normalized frequency. The high frequency region of **Fig. 3B** has discernably more area than the high frequency region of **Fig. 2B**. The power spectrum can be used to select components of high frequency regions and to determine the area or relative area under a high bandpass region. This information can be used to

25   determine threshold concentrations.

The first data set can also be filtered and back-transformed to produce a smoothed data set, which can then be subtracted from the first data set to produce a residual plot. **Figs. 2C** and **3C** show plots of the residuals of the filtered back-transformed data subtracted from the first data set. The residuals in **Fig. 3C** are larger than those in

30   **Fig. 2C**. The largest residual maximum in **Fig. 3C** is at the site of the paraprotein response. These plots can be used to determine the locations of any paraprotein regions above the predetermined threshold.

13

Figs. **2D** and **3D** depict the comparisons of the filtered back-transformed data and the first data set with arrows marking the paraprotein regions detected at a concentration above the second threshold. The smoothed back-transformed data from the normal sample in **Fig. 2D** matches the original data quite closely, while discrepancies may be seen between the smoothed, back-transformed data and the original data from the abnormal sample in **Fig. 3D**.

Figs. **2A** and **3A** depict the final result. A paraprotein was correctly found for the abnormal sample (**Fig. 3A**) at a normalized mobility of -0.15.

The methods of this invention are of greatest interest for the analysis of biological samples, or the detection and/or quantifying of specific components in biological samples. Typical samples include whole blood, plasma, serum, urine and cerebrospinal fluid. Human serum is one of the most common samples in need of analysis.

Thus, in another aspect, the present invention provides a method of monitoring paraprotein production in an individual, comprising;

(a)   subjecting an individual's first serum protein sample to capillary electrophoresis and detecting paraproteins at a first level using the method described above;

(b)   subjecting said individual's second serum protein sample to capillary electrophoresis and detecting paraproteins at a second level according to methods above; and

(c)   comparing the first level and the second level to monitor level of paraprotein production in said individual.

In this aspect of the invention, the comparing can be carried out by a skilled clinician or by computer programs which provide comparison routines, and the calculation of areas from a specific region, such as, for example, CDM 2.0 software, available from Bio-Rad Laboratories.

14

The following example is offered for purposes of illustration only.

EXAMPLE 1

Four patient sera samples were obtained from a university hospital in a midwestern state. Each of the samples contained a paraprotein, quantified at the hospital
5   using the Beckman Paragon gel at a level above 1.0 g/dL. Two were IgG, one was IgA and the other was IgM.

Each sample was diluted with a pool of normal sera to obtain paraprotein concentrations of 0.5, 0.2, 0.1 and 0.05 g/dL. The samples containing paraproteins at 0.1 g/dL were each run ten times. The neat patient sample and all dilutions were run in
10   duplicate. The normal pool was run four times along with each patient set, twice before and twice after the patient samples. A positive result was indicated by the presence of a "check gamma" message in the sample report.

Positive results were obtained on all the neat, 0.5 and 0.2 g/dL solutions. At the 0.1 g/dL level, 39 of the 40 runs evaluated indicated a positive result. At the 0.05
15   g/dL level, three of the eight injections had positive results. For the control, only one of 16 injections indicated a positive result (see Table 1).

TABLE 1

Concentration of Paraproteins	n	Positive*	Negative
> 1.0 g/dL	8	8	0
0.5 g/dL	8	8	0
0.2 g/dL	8	8	0
0.1 g/dL	40	39	1
0.05 g/dL	8	3	5
Negative pool	16	1	15
*Positive indicated by "Check Gamma" message on report.			

15

The foregoing is offered primarily for purposes of illustration. It will be readily apparent to those skilled in the art that the operating conditions, materials, procedural steps and other parameters of the systems and methods described herein may be further modified or substituted in various ways without departing from the spirit and scope
5   of the invention.

16

WHAT IS CLAIMED IS:

1.   A method of detecting the presence or absence of paraproteins in a serum protein sample, said method comprising;

(a)   separating components in said sample using capillary electrophoresis to generate a first data set;

(b)   subjecting at least a portion of said first data set to Fourier Analysis to generate a forward-transformed data set;

(c)   selecting any forward-transformed data sets having high frequency components above a first preselected threshold;

(d)   filtering and back-transforming data sets selected in step (c) to provide filtered, back-transformed data sets;

(e)   identifying the magnitude and location of residual maxima in said filtered, back-transformed data sets; and

(f)   comparing the location of any residual maxima having a magnitude above a second preselected threshold to a corresponding location in said first data set to detect the presence of paraproteins in said sample.

2.   A method in accordance with claim 1, wherein said first data set is a delimited mobility data set.

3.   A method in accordance with claim 1, wherein said first data set is a time data set.

4.   A method in accordance with claim 1, wherein said filtering uses filters selected from the group consisting of high-pass filters, low-pass filters, notch filters and combinations thereof.

5.   A method in accordance with claim 1, wherein said filtering uses high-pass filters and low-pass filters.

SUBSTITUTE SHEET (RULE 26)

17

6.   A method in accordance with claim 5, wherein said filter is matched to the frequency profile of potential paraprotein regions.

7.   A method in accordance with claim 1, wherein said paraproteins are detected at a level above about 50 mg/dL.

8.   A method in accordance with claim 1, wherein said paraproteins are detected at a level above about 100 mg/dL.

9.   A method in accordance with claim 1, wherein said paraproteins are detected at a level above about 250 mg/dL.

10.   A method in accordance with claim 1, wherein said first data set is a delimited mobility data set which is mobility zero corrected and normalized.

11.   A method in accordance with claim 1, wherein each of steps (a)-(f) are performed by automated instrumentation governed by computer software.

12.   A method of monitoring paraprotein production in an individual, said method comprising;

(a)   subjecting an individual's first serum protein sample to capillary electrophoresis and detecting paraproteins at a first level using the method of claim 1;

(b)   subjecting said individual's second serum protein sample to capillary electrophoresis and detecting paraproteins at a second level according to method of claim 1; and

(c)   comparing said first level and said second level to monitor level of paraprotein production in said individual.

13.   A method in accordance with claim 12, wherein said serum protein sample is human serum protein.

Fig. 1a. IgG PARAPROTEIN (2.2 g/dL)
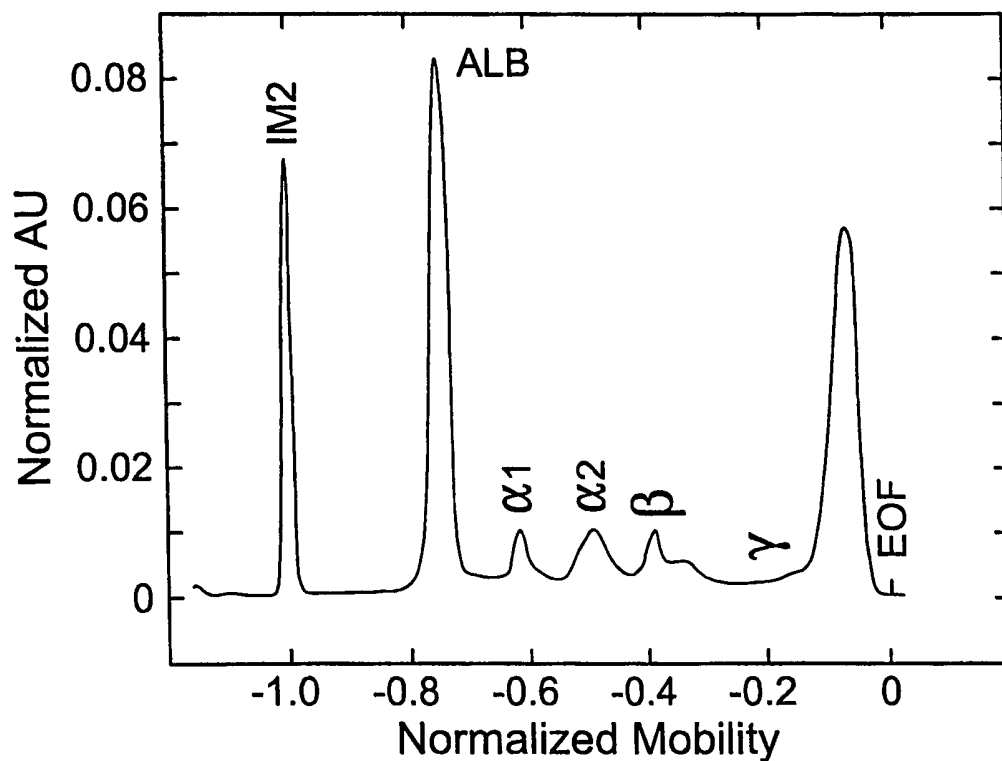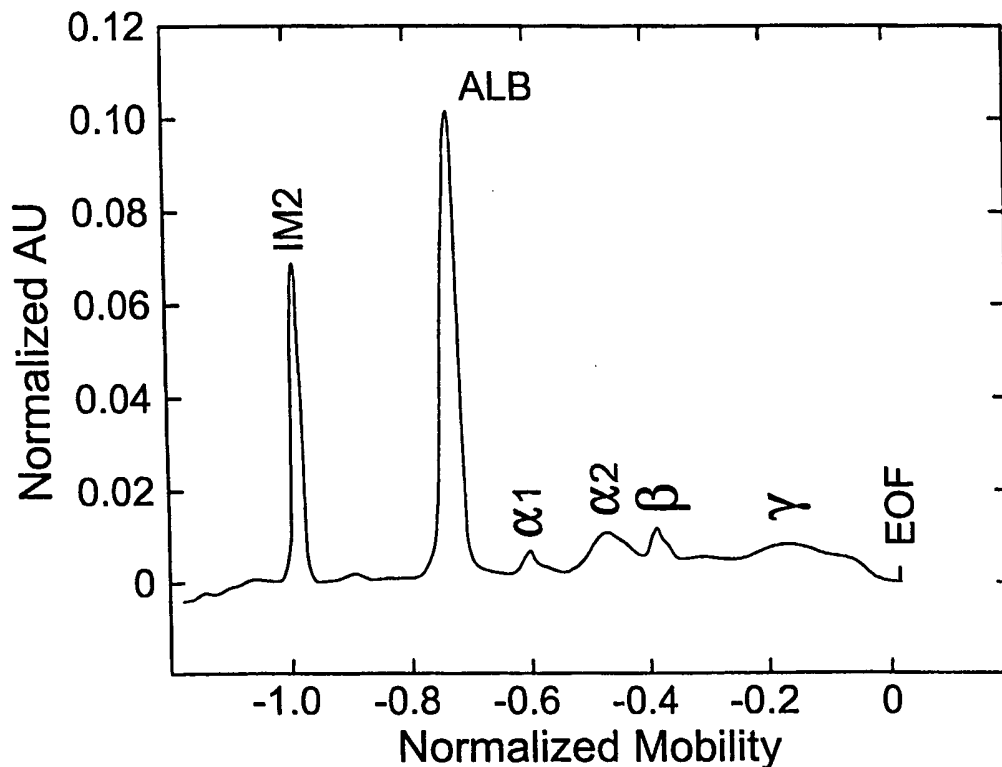

Fig. 1b. IgG PARAPROTEIN (Diluted to 0.1 g/dL)

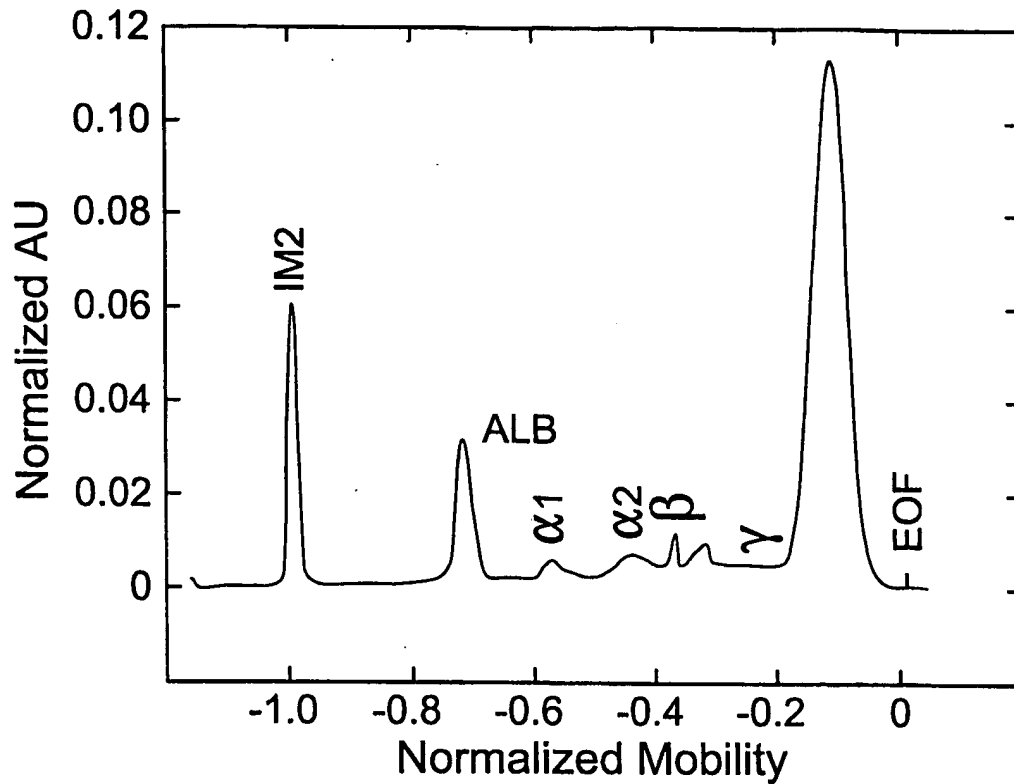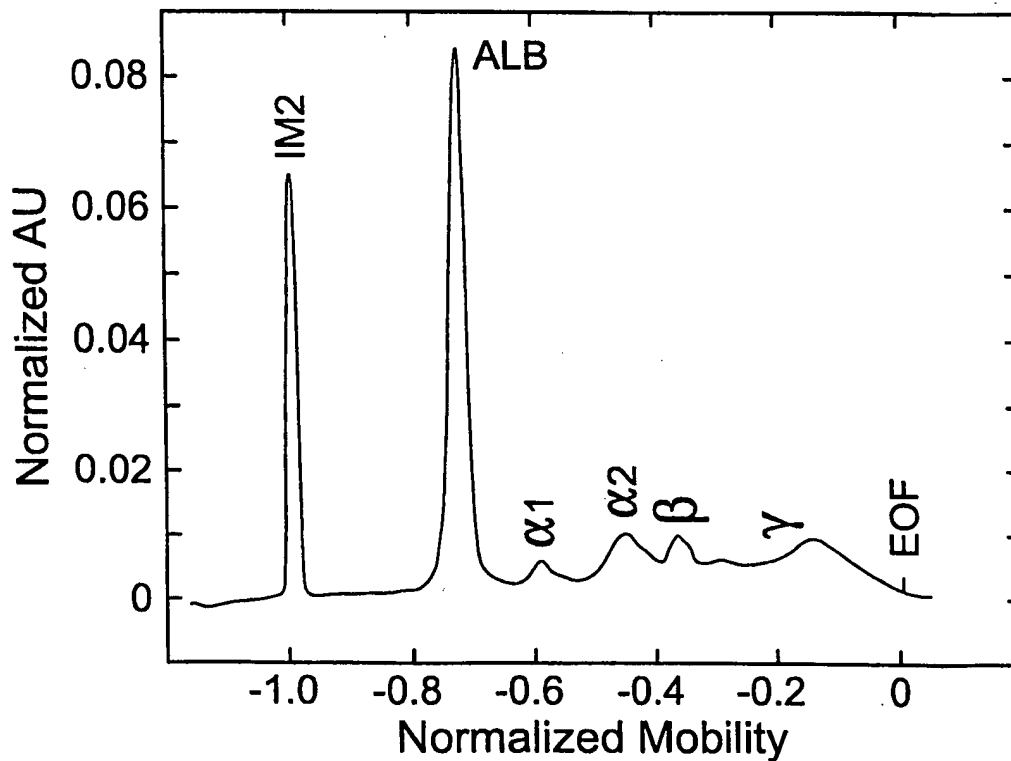## Fig. 1c. IgG PARAPROTEIN (7.1 g/dL)

Fig. 1d. IgG PARAPROTEIN (Diluted to 0.1 g/dL)
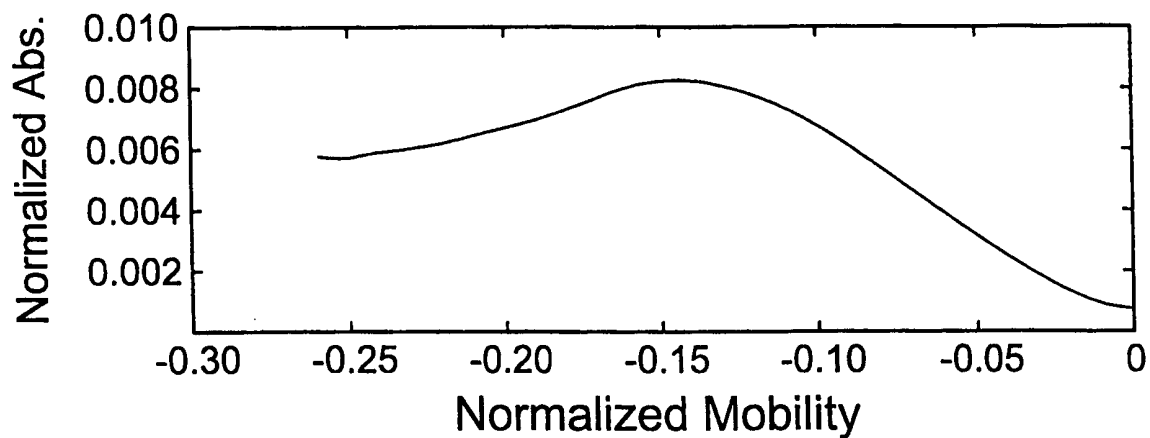
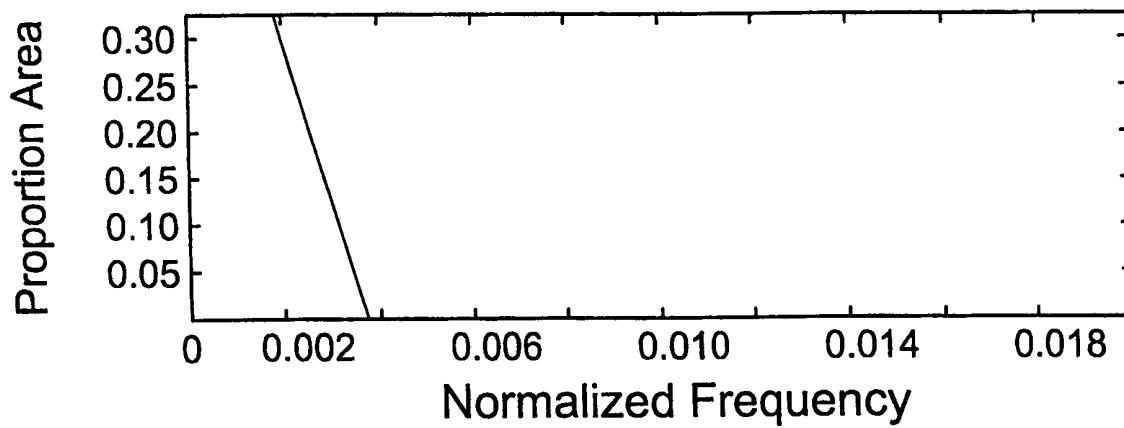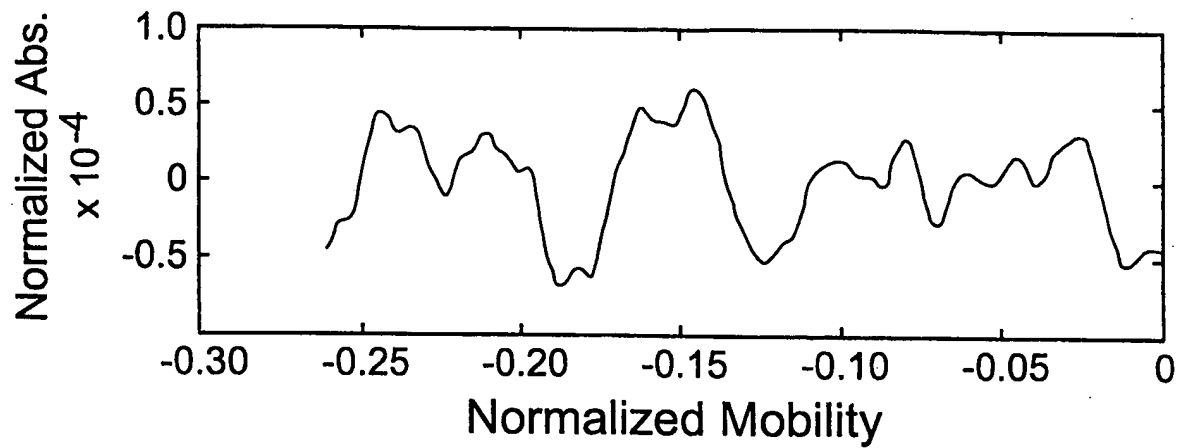*Fig. 2a.* Normal Serum

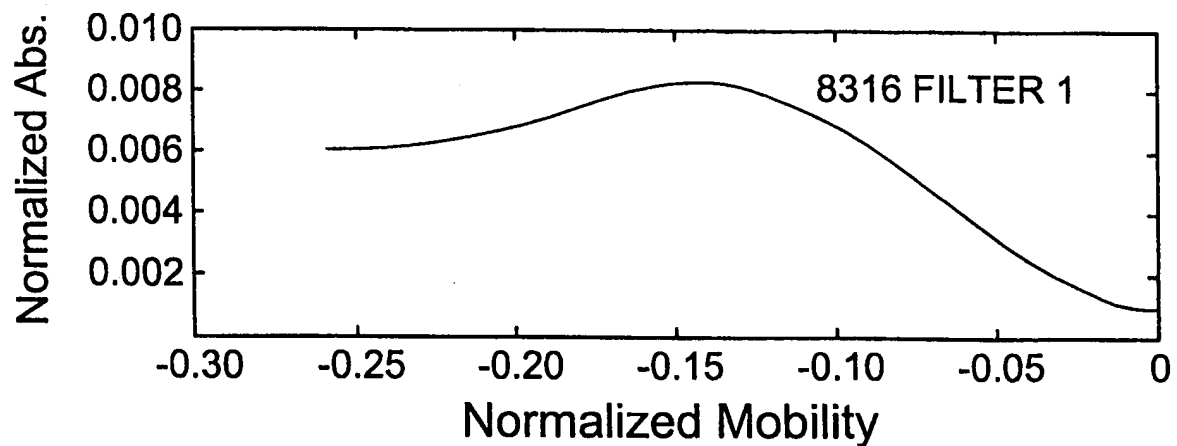*Fig. 2b.* Normal Serum

Fig. 2c. Normal Serum

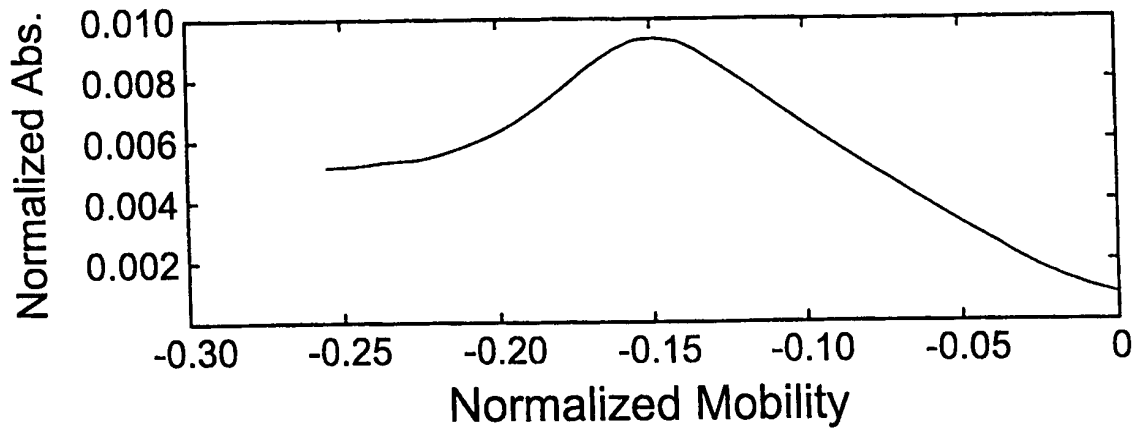*Fig. 2d.* Normal Serum

Fig. 3a. ·IGG Diluted to 0.1 g/dL
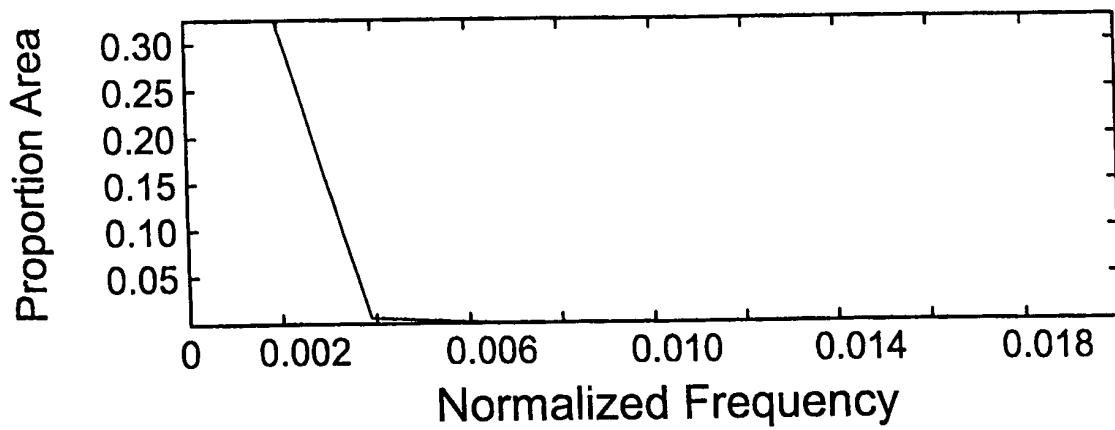

Fig. 3b. IGG Diluted to 0.1 g/dL
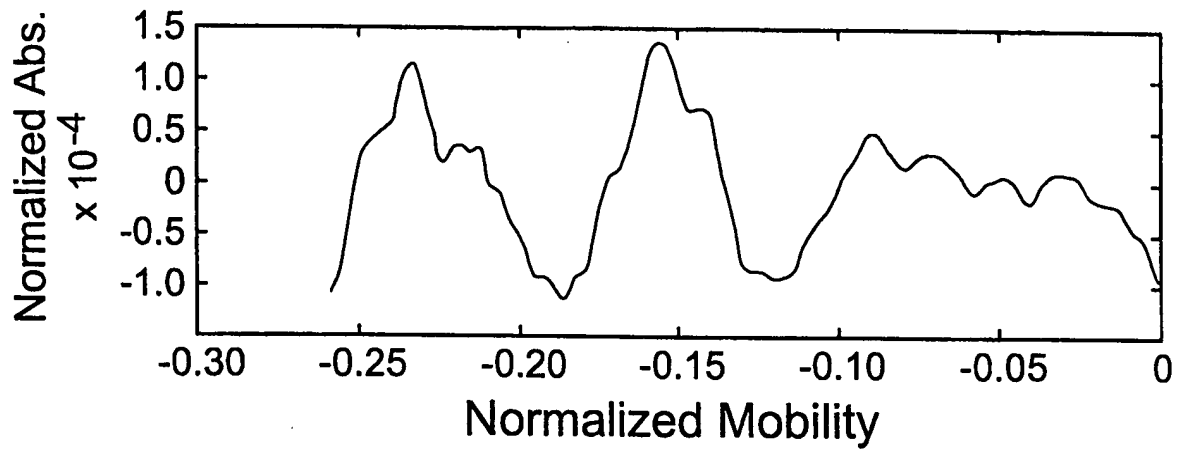
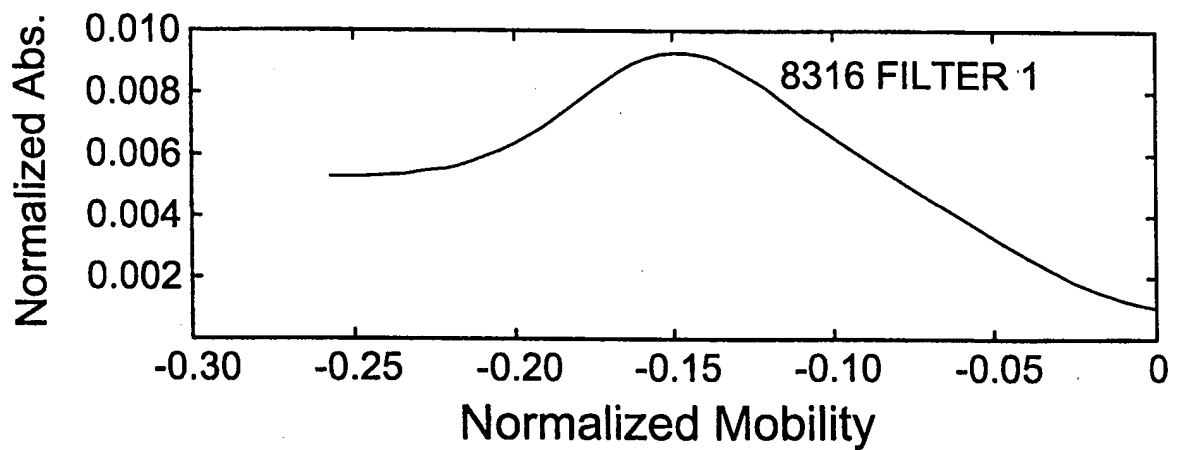*Fig. 3c.* IGG Diluted to 0.1 g/dL

Fig. 3d. IGG Diluted to 0.1 g/dL

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US98/14740

A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) : G01N 27/26
US CL : 204/452, 461

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 204/452, 461, 451; 702/22, 23, 32, 25; 436/516

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

WPIDS, JAPIO, CAPLUS, APS
search terms: Fourier, chromatograph?, deconvolut?, resolv?, peak, overlap, electrophore?

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	English language translation of JP 63117257 A (SAITO) 21 May 1988 (21/05/88), see pages 5 and 6; and Figures 1 and 2	1-13
Y	LASTERS et al. Background estimation in one-dimensional electropherograms of whole-cell protein extracts, 1983, 6, 508-511, see abstract; column 2 on page 508; "Corrections for the extrinsic part can be performed by the application of appropriate algorithms; e.g. low pass filtering ..." in column 1, page 508; column 1, first paragraph of page 509; column 2, page 508.	1-13
Y	US 5,567,282 A (WANG ET AL) 22 October 1996 (22/10/96), see abstract; column 1, line 32 - col. 2, ln. 42; col. 2, ln. 42; and col. 11, ln. 43-45.	7-9 and 13

☒ Further documents are listed in the continuation of Box C.   ☐ See patent family annex.

*	Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A"	document defining the general state of the art which is not considered to be of particular relevance		
"B"	earlier document published on or after the international filing date	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"O"	document referring to an oral disclosure, use, exhibition or other means		
"P"	document published prior to the international filing date but later than the priority date claimed	"&"	document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
25 NOVEMBER 1998	**09 DEC 1998**
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231	Authorized officer ALEXANDER NOGUEROLA
Facsimile No. (703) 305-3230	Telephone No. (703) 308-0661

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 5,490,909 A (WANG ET AL) 13 February 1996 (13/02/96), see column 1, lines 15-37; col. 8, ln. 45-53; col. 11, ln. 39 - 53; and col. 1, ln. 15 - col. 2, ln. 2.	7-9 and 13